Defense Advanced Research Projects Agency DARPA

A DARPA Information Processing Technology Renaissance: Developing Cognitive Systems

Ron Brachman

Zach Lemnios

Information Processing Technology Office

Defense Advanced Research Projects Agency

INFORMATION PROCESSING TECHNOLOGY OFFICE

DARPA/IPTO and the Computing Revolution

- DARPA is credited with "between a third and a half of all the major innovations in computer science and technology" – Michael Dertouzos, *What Will Be* (1997)
- The information technology revolution of the second half of the 20th century was largely driven by IPTO
  - Time-sharing, interactive computing, personal computing
  - ILLIAC IV
  - ARPANET, Internet
- J.C.R. Licklider (first IPTO Director) had the goal of human-computer symbiosis
  - The computer as communications mediator
  - Agents ("OLIVERS")
  - Intergalactic computer network

Now: A chance to reclaim this legacy and help orchestrate the next revolution

A Problem of National Importance

- Computer systems are the critical backbone of DoD systems and the national infrastructure
  - Virtually all important transactions involve massive amounts of software and multiple computer networks
- While *computational performance* is increasing, productivity and effectiveness are not keeping up – in fact, system complexity may actually be reversing the information revolution
  - The cost of building and maintaining systems is growing out of control
  - Systems have short lifespans with decreasing ROI
  - Demands on expertise of users are constantly increasing
  - Users have to adapt to system interfaces, rather than vice versa
- As a result, systems have grown more rigid, more fragile, and increasingly vulnerable to attack
  - Ultimate asymmetric threat: one person could destroy significant national infrastructure
  - We need to change the game to achieve an urgent and necessary quantum leap in capability and productivity

Our Solution

Developing Cognitive Systems:

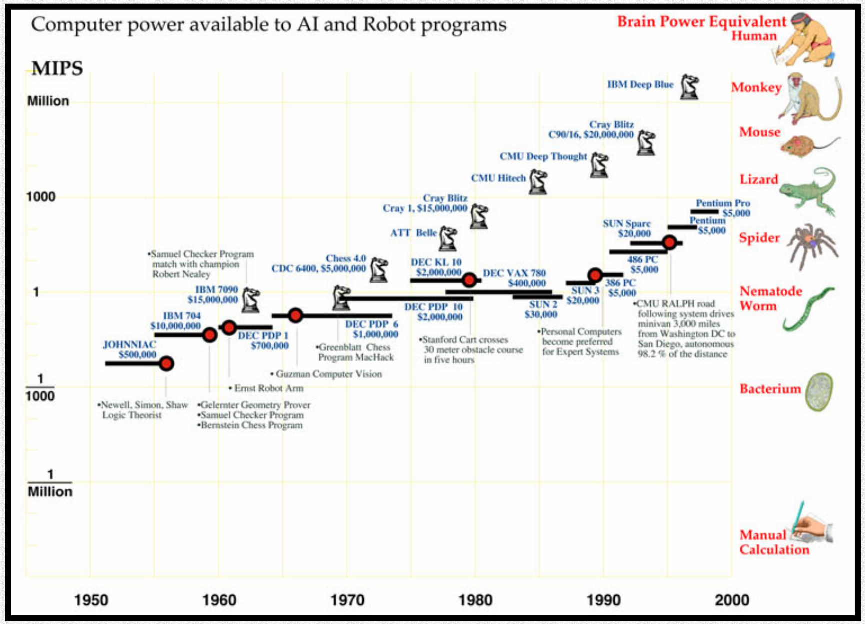
Systems that know what they're doing

- A cognitive system is one that
  - can reason, using substantial amounts of appropriately represented knowledge
  - can learn from its experience so that it performs better tomorrow than it did today
  - can explain itself and be told what to do
  - can be aware of its own capabilities and reflect on its own behavior
  - can respond robustly to surprise

Systems that know what they’re doing can…

- …reflect on what goes wrong when an anomaly occurs and anticipate its occurrence in the future
- …assist in their own debugging
- …reconfigure themselves in response to environmental changes
- …respond to naturally-expressed user directives to change behavior or increase functionality
- …be configured and maintained by non-experts
- …thwart adversarial systems that don’t know what they’re doing
- …last much longer than current systems

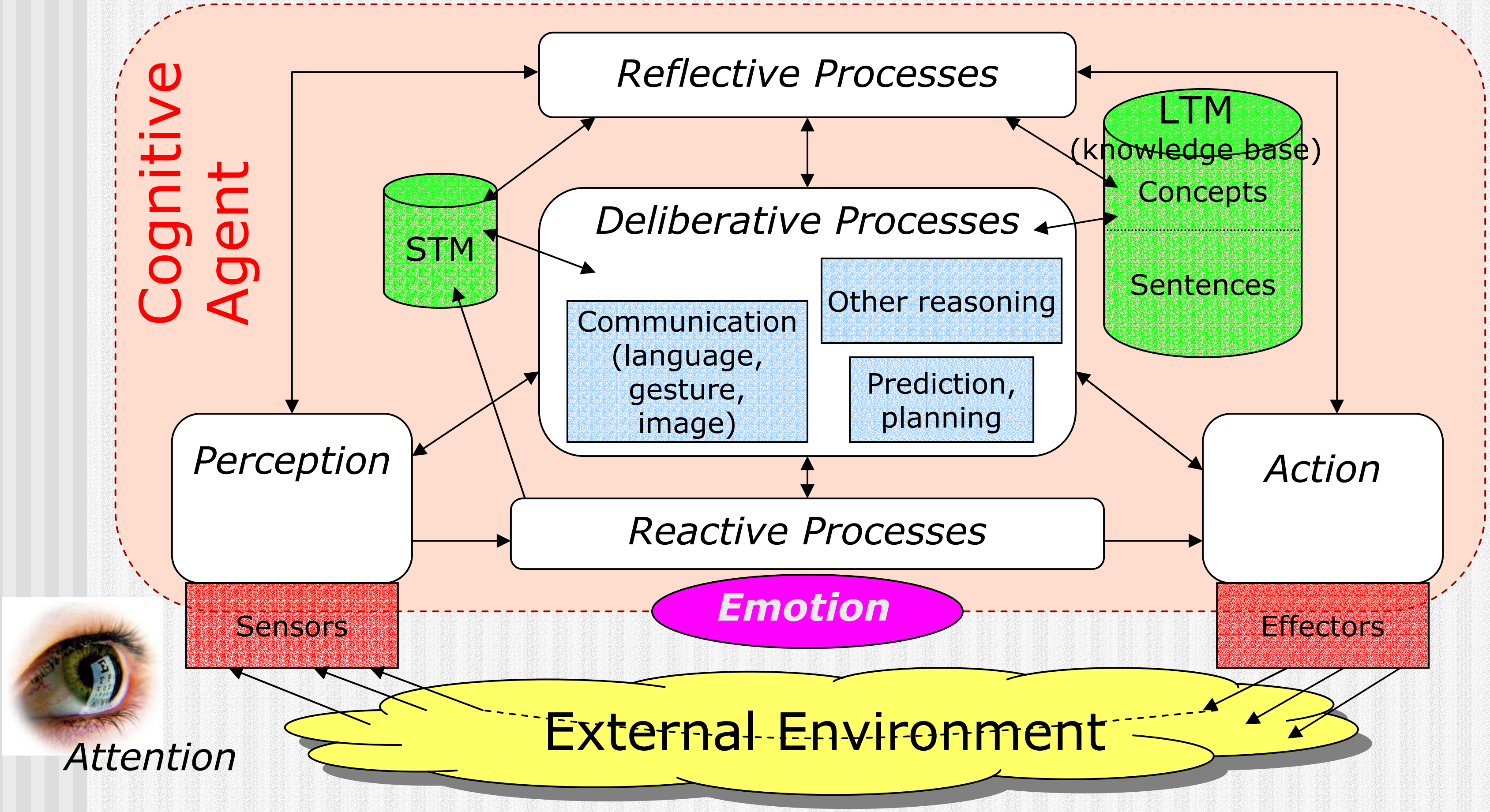
Why Now?



(H. Moravec, CMU)

- Human-level scaling of HW technology is on the horizon
- Advances in understanding of human neural systems
- Cognitive technology (from AI and elsewhere) is working in bits and pieces, ranging from large-scale knowledge bases to machine learning in support of data mining

Rough Anatomy of a Cognitive Agent



Notes on Architecture

- Long-/short-term memory (LTM/STM) use knowledge representation
  - Knowledge base has many components: concepts, facts, rules of thumb, people, smells, …
- Different types of learning expected in different components (e.g., learned reactions, learned facts, learned concepts, learned problem-solving strategies)
- Reflective component may distinguish between simple reflection (observation) and "self"-oriented reflection (consciousness?)
- "Other reasoning" includes comparing, plan recognition, analogy, envisioning, etc.
- Humans cannot reliably inspect their own processes, but it may be productive to allow an artifact to do so

- *Key questions:*

  What's missing? Is the strawman architecture adequate to do the job? Do we need a radical change in our view of the architecture to make a big difference?

Cognitive System Examples

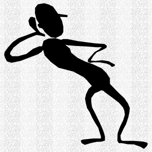
Self-aware software

- Extend functionality by means of interactive dialogue
- Actively assist programmer in debugging

Adaptive, cognitive networks

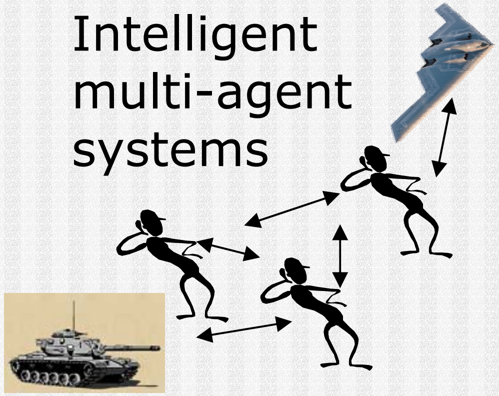
- Explain causes of network delays
- Self-reconfigure by reasoning about traffic, anomalies
- Learn and adapt to new attacks

Perceptive, instructable agents



- Personalize via learning
- Reconfigure by natural language request ("what do you want me to become?")
- Perceive important threads in large amounts of data

Intelligent multi-agent systems

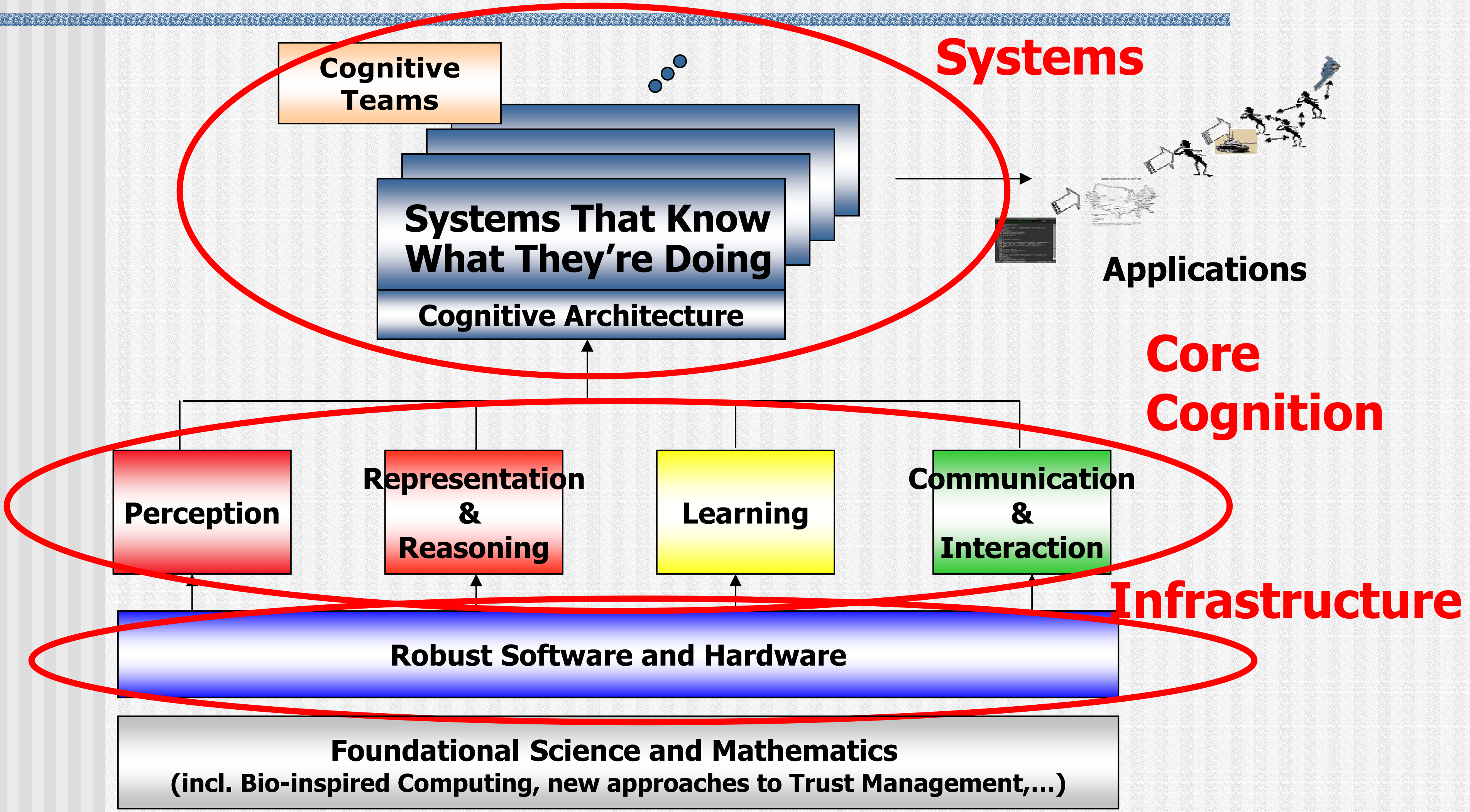


- Automatic cross-component coordination, with shared goals
- Overall cost minimization

Teams of Cognitive Systems

- It is not sufficient to create technology for individual cognitive agents
  - Agents will need to interact with other agents, humans, and non-cognitive systems
  - Coordination and communication are essential – but because of autonomy and cognition (including planning, counter-planning, and possible deceit), the issues are much more complex than with earlier generations of computing systems
  - Entire systems can take on goals that individual agents cannot achieve themselves

Cognitive Systems Thrusts



Application Foundations
Some Key Functional Capabilities

- “Needles and threads”/Perceptive agents
  - ability to detect important small probability events and chain together key observations – at scale
- “Form-fitting” interfaces/Communications assemblers
  - instructable and adaptable
- Strategic envisioning
  - computational imagination for scenario planning, assessment of plausible outcomes, prediction of next steps
- National Knowledge Bank
  - a knowledge bank of critical assets and know-how for broad use in DoD applications
- Adaptive networks
  - capable of detecting threats and automatically responding
  - testbed for distributed cognitive capabilities

Initial Challenge Context

- Persistent, personal partner/associate systems
  - Learn from experience
  - Learn what you like and how you operate
    - by observation
    - by direct instruction or guidance, in a natural way
  - Imagine possible futures, anticipate problems and needs
  - Omnipresent/always available
- Examples
  - Commander’s (C^2) assistant
  - (Intelligence) Analyst’s associate
  - Personal executive assistant/secretary
  - Disaster response captain’s “RAP” (robot/agent/person) team

Initial Challenge Context

An Enduring, Personalized, Cognitive Assistant

Radar O’Reilly
- observed
- anticipated
- planned
- worked autonomously (but supervised)

An Enduring, Personalized, Cognitive Assistant

- Will have and use *knowledge* of the domain, task
- *Cognitive awareness*: will have experiences; perceptual input integrates with knowledge, model-based filtering
- Can *imagine* possible futures
- Can *decide* what to do and act in real time (prioritize)

An Enduring, Personalized, Cognitive Assistant

- Will have and use *knowledge* of the domain, task
- *Cognitive awareness*: will have experiences; perceptual input integrates with knowledge, model-based filtering
- Can *imagine* possible futures
- Can *decide* what to do and act in real time (prioritize)
- *Learns,* including by observing partner
- Can be *advised* and *guided,* and can *explain*

An Enduring, Personalized, Cognitive Assistant

- Will have and use *knowledge* of the domain, task
- *Cognitive awareness*: will have experiences; perceptual input integrates with knowledge, model-based filtering
- Can *imagine* possible futures
- Can *decide* what to do and act in real time (prioritize)
- *Learns,* including by observing partner
- Can be *advised* and *guided,* and can *explain*
- Must know how to *cooperate* (be a team player)
- Uses *multi-modal*, broad-spectrum interaction
- Should be available everywhere - *omnipresent*
- Must be *trustworthy*

An Enduring, Personalized, Cognitive Assistant

- Will have and use *knowledge* of the domain, task
- *Cognitive awareness*: will have experiences; perceptual input integrates with knowledge, model-based filtering
- Can *imagine* possible futures
- Can *decide* what to do and act in real time (prioritize)
- *Learns,* including by observing partner
- Can be *advised* and *guided,* and can *explain*
- Must know how to *cooperate* (be a team player)
- Uses *multi-modal*, broad-spectrum interaction
- Should be available everywhere - *omnipresent*
- Must be *trustworthy*
- Must *learn continuously*
- Must be able to *survive*, operate through problems

An Enduring, Personalized, Cognitive Assistant

- Will have and use *knowledge* of the domain, task
- *Cognitive awareness*: will have experiences; perceptual input integrates with knowledge, model-based filtering
- Can *imagine* possible futures
- Can *decide* what to do and act in real time (prioritize)
- *Learns,* including by observing partner
- Can be *advised* and *guided,* and can *explain*
- Must know how to *cooperate* (be a team player)
- Uses *multi-modal*, broad-spectrum interaction
- Should be available everywhere - *omnipresent*
- Must be *trustworthy*
- Must *learn continuously*
- Must be able to *survive*, operate through problems

Research Challenges [1/5]

- Architecture
  - What is the most effective coupling between the perceptual and deliberative components of a cognitive system?
  - How can we build systems that can keep a watchful eye on themselves? How can you do effective reflection and still operate in real time?
  - How do we design and build associate systems with adjustable degrees of autonomy?
  - What role should affect/emotion play?
- Perception
  - Can we use insights from human and animal perception to help computers sort through overwhelming amounts of visual and auditory data in real time, and detect important, low-frequency events?
  - Can insights from neuroscience and elsewhere provide hints on how to help people deal with information overload?

Research Challenges, cont'd. [2/5]

- Representation and Reasoning
  - How can plausible but not logically sound reasoning be used in a consistent and pragmatic way to get to reasonable conclusions?
  - Can you build a real-world-scale cognitive system that has a principled foundation for its representation and reasoning capabilities?
  - How (if at all) might it be possible for a system to actively aid its creator in its own debugging?
  - How do we build systems that can deal with proper understanding and prioritization of standing orders, given complex and even conflicting goals?
  - Can an artificial system augment a human's capacity to imagine future scenarios and help prepare for never-before-encountered situations?

Research Challenges, cont'd. [3/5]

- Learning
  - How might a cognitive system learn the salient things from each experience it has and later use what it has learned to interpret and successfully cope with new situations?
  - How do you find the right remembered experiences to apply to each new situation?
  - Can we build systems, e.g., networks, intelligent enough to recognize intrusion attacks and then through experience learn how to repel subsequent attacks in a general way?
- Communication
  - Can we build an "instructable" interface that will accept direct naturally-expressed guidance as to the desires of its user? (what ever happened to McCarthy's "Advice Taker"?)
  - How might an intelligent system use context (of various sorts) to help disambiguate complex natural language expressions and other actions/events appropriately?

Research Challenges, cont'd. [4/5]

- Teams
  - Can we create teams of agents whose "collective IQ" is at least as high as the "IQ" of its components?
  - Can insights from cognitive team research help improve human group decision-making?
  - Are some types of team/collective structure more vulnerable to disruption than others? Are some more robust?
  - What notions of trust and accountability are critical to the operation of systems with artificial cognitive agents?

Research Challenges, cont'd. [5/5]

- Foundations and Infrastructure
  - Can insights from cognitive systems yield new general approaches to reliable, fault-tolerant, secure software and hardware systems?
  - Can insights from neuroscience lead to breakthroughs in the design and building of artificial cognitive systems?
  - Can novel hardware designs make a big difference in how we conceive of cognitive systems?
  - Can modern "post-PC" distributed computing fabrics facilitate building more robust intelligent systems?
  - Can Shannon's theory of "information" be extended to account for how information is formed and used in the "head" of a cognitive agent and transmitted to another?

IPTO Will Lead the Way

Building on a 40 year legacy of changing the world, IPTO will drive dramatic improvement in computing and revolutionary change in how people think of and use computational systems

…but it all depends on your ideas

…and our collective ability to deliver

Our office BAA (02-21):

http://www.eps.gov/spg/ODA/DARPA/CMO/BAA02-21/listing.html